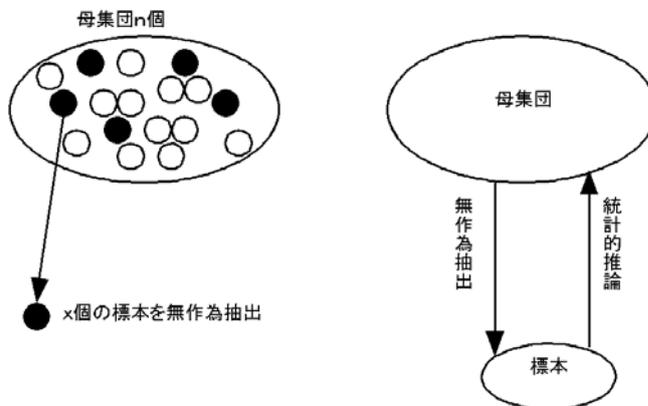


## 第4回 統計的推定

★ 教材「生物統計学\_統計的推定と検定 2013」を予習しながら空所を埋めておくこと

### A. 統計的推定と検定

母集団から無作為抽出した標本から母集団についてなんらかの推論を行う。この場合、統計から行う推論には統計的（ ）と統計的（ ）の2つがある。



推定 統計的に標本の統計量から母集団の母数（母平均，母標準偏差など）を推論することを統計的推定という。

例：視聴率調査を 200 人に対して行い，番組Aの視聴率を推定した。

検定 統計的に標本の統計量から母数に関する予想の真偽を検証することを統計的検定という。

例：視聴率調査を 200 人に対して行い，番組Aの視聴率が 20%以上あるのかを検定した。

例：A社とB社の車の排気ガスに含まれる窒素酸化物はA社の方が多いのかを検定した。

★ 次の文章は統計的推定か，統計的検定かを考えよ。（答え合わせは授業でします）

- ① A牧場の牛から 20 頭を選んで，乳脂肪率を測定し，A牧場の牛全体の乳脂肪率を知りたい。  
統計的推定 ・ 統計的検定
- ② 番組Bの視聴率を知りたい。  
統計的推定 ・ 統計的検定
- ③ 窒素施肥量  $10\text{kg}/\text{m}^2$  を与えると，コシヒカリの収量がいくらになるかを知りたい。  
統計的推定 ・ 統計的検定
- ④ A牧場の牛は去年から飼料をF社からG社に変えた。乳脂肪率が変化したかを知りたい。  
統計的推定 ・ 統計的検定
- ⑤ 番組Bは番組Cより視聴率が高いのかを知りたい。  
統計的推定 ・ 統計的検定
- ⑥ 1日当たりの喫煙本数が増えるほど，肺ガンの危険性が増えるのかを知りたい。  
統計的推定 ・ 統計的検定

次に統計的推定の実例を考えてみよう。

★ 教材「生物統計学\_\_二項分布を利用した統計的推定 2013」を予習しながら空所を埋めておくこと

## B. 二項分布, ポアソン分布を利用した統計的推定

### 1. 点推定

例1: さいころを 1200 回振る. このような実験を何回もやる場合, 平均すれば, 1 の出る回数は ( ) 回である.

例2: A さんが実験に使っている道具は何回か使用すると壊れてしまう. 1 回使用につき, 0.02 の確率で壊れる (使用回数によって確率は変わらないとする). A さんは毎日, この道具を 50 回使う. だから平均で道具は ( ) 個壊れる. そこで実験は 10 日間行うので, この道具をちょうど ( ) 個だけ購入した.

しかしながら, このような点推定では推定の不正確さを示していないから困った問題が起こる.

例1の場合, 確かに平均すれば 200 回だけ 1 がでるかもしれないけれども, 実際にちょうど 200 回だけ 1 がでる確率は二項分布で計算すると ( ) となり, めったには起こらない. 例2の場合, 10 日間のうちに 2 個より多く道具が壊れる可能性もある. たしかに 10 個ちょうど壊れる確率は ( ) と最大であるけれども, 11 個以上壊れる確率は ( ) もある. しかし, だからといってやみくもに道具を買おうと費用も置く場所にも困る.

### 2. 区間推定

一般に確率 95%あるいは 99% (この確率を信頼率と呼ぶ) のもとで, 平均など示したい値の範囲 (この範囲を信頼区間と呼ぶ) を示すのが一般的である.

例1: さいころを 1200 回振る. 二項分布によって, このような実験を何回もやる場合, 95%の確率で 1 の出る回数は ( ) 回の間に入る. 99%の確率で ( ) 回の間に入る.

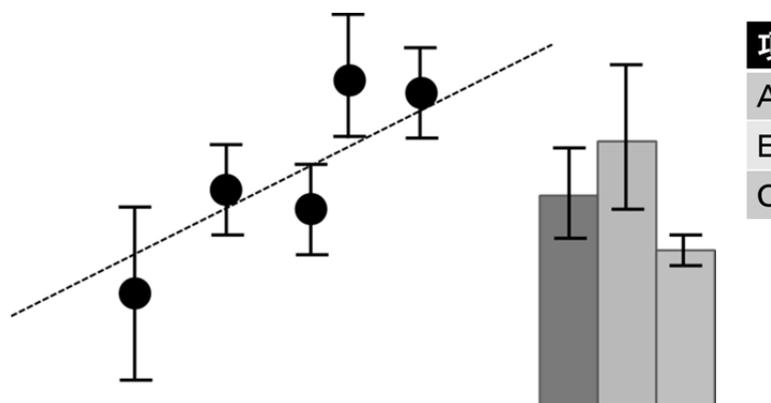
例2: A さんが実験に使っている道具は何回か使用すると壊れてしまう. 1 回使用につき, 0.02 の確率で壊れる (使用回数によって確率は変わらないとする). A さんは毎日, この道具を 50 回使う. だから平均で道具は 1 日に 1 個壊れる. 実験は 10 日間行う. 二項分布から, この道具が 10 日間のうちに壊れる個数は 95%の確率で ( ) 個, 99%の確率で ( ) 個の範囲に入る. そこで ( ) 個購入した.

実際には道具が壊れる確率は二項分布には従わない. しかし, 生存関数といって, 生物の寿命や器具・機械の寿命に関する分布があり, それを利用することも可能である. ラーメン屋に並ぶ行列などもある仮定の下に確率分布を考えることができ, それを待ち行列関数という.

## 信頼区間の例

台風の子報円 信頼率 70%の確率で暴風域圏内に入る

誤差 個々のデータが持っている誤差をグラフ上では誤差線（エラーバー）で、表では±標準偏差あるいは標準誤差という形式で表す。（信頼区間を示すとは限らないので注意）



項目	処理あり	処理なし
A	23.8±0.6	18.4±0.0
B	12.3±1.4	10.8±0.7
C	46.0±2.1	61.5±4.0

何パーセントの確率（信頼率）の信頼区間とすべきかは目的によって変わる

台風の子報円ではあまりに信頼率を高くしてもかえってわかりにくくなる

震度 6 強の地震で家が倒壊する恐れのある範囲ならどうだろうか？

自動車、宇宙船などの部品の信頼率だったらどのくらいがよいだろうか？

### ★ おまけ

信頼区間で考えると・・・ ある試験には問題が 10 問あり、各問に答えが 5 つ用意されている。わからない問題にはランダムに答えるとする。ある生徒は 4 問正解した。真の実力（すなわち実際に正解を知っていたのは何問だったか）を計算してみよう。（確率・統計入門 小針峴宏 岩波書店）

ひとつもわからなかった確率 0.074

ひとつだけわかっていた確率 0.148

2 つわかっていた確率 0.247

3 つわかっていた確率 0.309

4 つわかっていた確率 0.221

信頼区間をつけて区間推定するなら 77.7%の信頼率で 2~4 問わかっていたとなるのだ

さて問題数の少ない選択式問題は実力を適切に評価するだろうか？

例えば、たった 5 問の三択問題を試験でやればどうなるだろうか。

厳密そうに見える選択式試験も意外とあいまいなところもあるのだ！

授業での演習問題

ある市では1日平均で3回、救急車を発進している。もし救急車の発進数がポアソン分布に従うとすれば、99.9%以上の確率で救急車の不足が起きないようにするには救急車を何台、用意すればよいか？

★ 教材「生物統計学\_\_標準誤差 2013」を予習しながら空所を埋めておくこと

C. 標準誤差

1. 推定した平均の精度

例：宍道湖に生育するある魚が今回の大水害でどのような影響を受けたかを明らかにするために、魚を捕獲して調査しようとした。

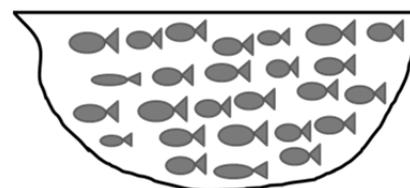
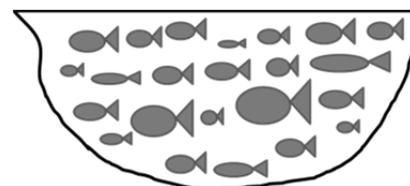
S 県は宍道湖に生息するある魚の平均体重を調査することを3つの会社に依頼した。

A 社は3匹で、B 社は20匹で、C 社は100匹で 平均体重を推定しようとした

平均をいちばん精度よく推定する会社はどれか？ 答え（ ）

右のような2つの池からそれぞれ100匹ずつ魚を捕獲して調査した

どちらの池での調査の方が平均の精度が高いだろうか？ 答え（ ）



標本から母集団の平均を推定する場合、

① 標本の数を多くするほど、

母集団の平均を推定する精度は

( 高くなる ・ 変わらない ・ 低くなる)

② 標本の数が同じであれば、母集団の分散が大きいほど、

母集団の平均を推定する精度は

( 高くなる ・ 変わらない ・ 低くなる)

では推定精度を定量的に評価できないのか？

## 2. 標準誤差の計算

標本数  $n$ ，標準偏差  $SD$  のとき，標準誤差  $SE$  は

$$SE = \frac{SD}{\sqrt{n}} \quad \text{となる.}$$

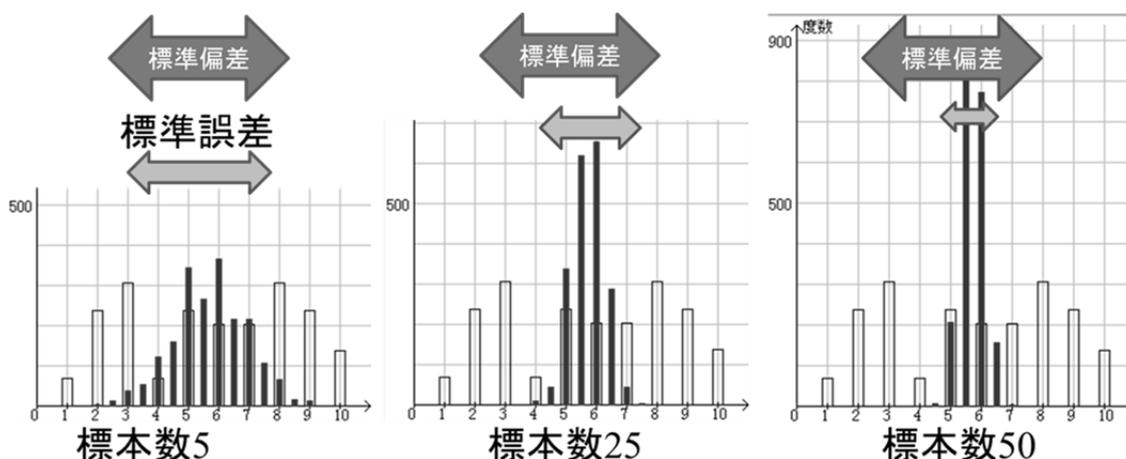
例：ある鳥の 100 個の卵の重さを量ったところ，平均 25.3g，標準偏差 3.0g であった．したがっ

て，標準誤差  $SE = \frac{SD}{\sqrt{n}} = \frac{3.0}{\sqrt{100}} = 0.3g$  である．

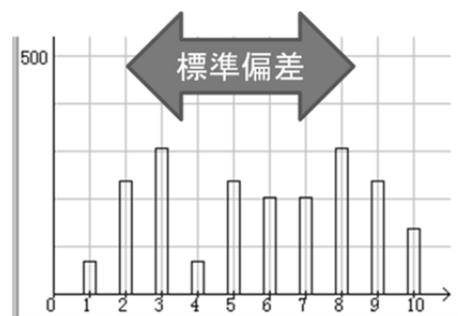
例：ある鳥の卵 10 個の重さを量ったところ，15.4, 15.8, 15.9, 16.1, 16.2, 16.4, 16.6, 16.8, 16.9, 17.5g であった．平均は 16.4g，標準偏差は 0.6g，標準誤差は 0.2g である．

16.4±0.2 (mean±SE) と論文などでは表記される (SE は standard error : 標準誤差)

## 3. 標準誤差の意味



標準誤差とは標本平均から母集団平均を推定するときどれくらい精度があるかを示す指標である．少ない標本から得た標本平均は母集団平均を推定するには誤差が大きい (上左図)．しかし標本数を増やせば精度が高まる (上右図)．



#### 4. 標準誤差の実際の計算

例：ある鳥の100個の卵の重さを量ったところ、平均25.3g、標準偏差3.0gであった。したがって、標準誤差  $SE = \frac{SD}{\sqrt{n}} = \frac{3.0}{\sqrt{100}} = 0.3g$  である。

$$SE = \frac{SD}{\sqrt{n}} = \frac{3.0}{\sqrt{100}} = 0.3g$$

=3/SQRT(100)	0.3
--------------	-----

エクセルによる標準誤差の計算

集めてきたデータそのものから、標準誤差をエクセルで計算するには標準偏差、サンプル数をそれぞれ関数から計算する方法と分析ツールにある基本統計量から計算する方法がある。さらに100個以内のデータであれば計算するシートを第2回基本統計量タブのところに作ってある。

	エクセルでの計算式	答え
平均	=AVERAGE(C12:C21)	16.36
標準偏差	=STDEV(C12:C21)	0.613188
サンプル数	=COUNT(C12:C21)	10
標準誤差	=STDEV(C12:C21)/SQRT(COUNT(C12:C21))	0.193907
列1		
平均		16.36
標準誤差		0.193907
中央値 (メジアン)		16.3
最頻値 (モード)		#N/A
標準偏差		0.613188
分散		0.376
尖度		0.00258
歪度		0.336574
範囲		2.1
最小		15.4
最大		17.5
合計		163.6
標本数		10

下の測定値にデータ(100個以内)を入れると平均、分散、標準偏差、メジアン、レンジ、変動係数を自動的に計算する。ただし測定値に値を誤入力したときはDelキーで削除すること。セルを移動させると式が変わってしまう。

標本番号	測定値		
1		平均	#DIV/0!
2		分散	#DIV/0!
3		標準偏差	#DIV/0!
4		メジアン	#NUM!
5		レンジ	0
6		変動係数	#DIV/0!
7		標準誤差	#DIV/0!
8			
9		標準誤差は第6回で学ぶ	
10			

#### 予習での練習問題

① 宍道湖で50匹のコイの標本を得た。平均体重3.5kg、標準偏差1.2kgだった。この標本の標準誤差を計算せよ。

② 宍道湖で10匹のコイの標本を得た。それぞれの体重は1.3, 1.8, 2.3, 2.9, 3.5, 3.6, 3.8, 4.0, 4.3, 5.1kgであった。この標本の標準誤差を計算せよ。

標本の標準偏差は標本数を増やしてもほとんど変わらない。一方、標準誤差は標本数を増やしていくと次第に減っていく。このことは湖のコイをいくらたくさん取ったとしても、湖のコイの体重そのもののばらつき（分散、標準偏差）が大きくなったり、小さくなったりすることはないことから、標本の標準偏差そのものが標本数には関係しないことが理解できるだろう。一方、コイをたくさん取れば、池のコイの体重の平均についてはより信頼の得られる値を推定できるだろう。このような信頼の程度を示すのが、標準誤差である。標本を増やせば、標準誤差が小さくなることが理解できるだろう。

### 授業での演習問題

S君は大学の研究で近くの森を調査している。そこに生息するある生き物の平均体重を5%以内の精度で推定したい。5%の精度とはここでは精度(%) = 標準誤差 ÷ 平均 × 100%とする。

この生き物の体重について、母平均は250g、母標準偏差は100gとわかっている。すると何匹、調査すれば精度を5%以下にできるだろうか？

★ 教材「生物統計学\_\_母平均に関する統計的推定 2013」を予習しながら空所を埋めておくこと

#### D. 平均に関する統計的推定

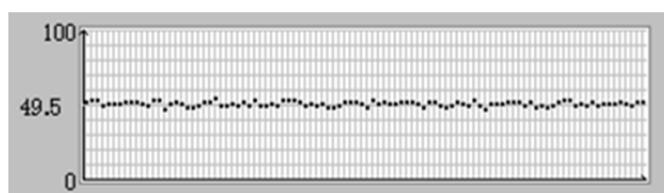
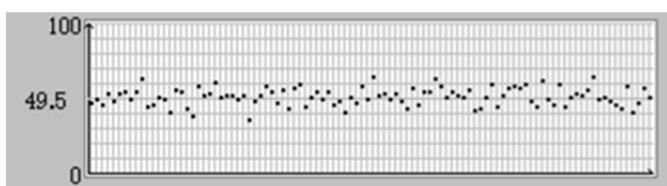
##### 1. 点推定 1つの値で母数を推定する

例：朝酌川で無作為に数地点選び、鯉を釣った。鯉の体重の標本平均 5.0kg なので、朝酌川の鯉の平均体重を 5.0kg と推定した。

しかし、点推定では、推定値がどのくらい信頼できるかを示すことができない。

たった5匹の標本で点推定するのはかなり危険そう

しかし、いつも何百匹も調査できるわけでもない(5匹でもどのくらい精度があるのかを示したい)

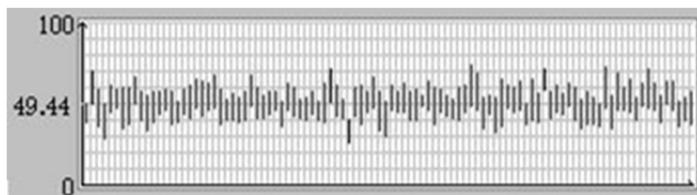


2. 区間推定 母数がある幅を持つ信頼区間とともに推定する.

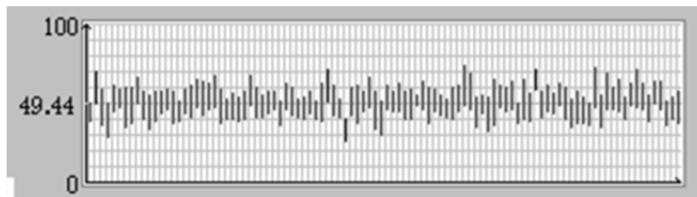
例：朝酌川で無作為に数地点選び，鯉を釣った．サンプル数 100 匹の鯉の体重の標本平均 5.0kg，標準偏差 2.0kg なので，朝酌川の鯉の平均体重は 95%の信頼区間をつけて， $4.6 \leq \mu \leq 5.4kg$  と推定した．このとき，母集団（朝酌川のすべての鯉）の母平均は 95%の確率で  $4.6 \leq \mu \leq 5.4kg$  の範囲に入る．

信頼率は統計量を使用する目的に応じて決定する．95%信頼区間の場合，100 回のうち 5 回，区間推定の値に母集団の平均が含まれない（すなわち推定を誤る）ということが起こる．信頼率を高める，例えば 100%にしたらいのではないかと考える人もいるかもしれないが，そのときには信頼区間はきわめて大きくなり，役に立たない．例えば，視聴率の 100%信頼区間は 0~100% だというようなもので，疑いはできないが，この数値では何の役にも立たない．信頼率は必要に応じた大きさを設定する．

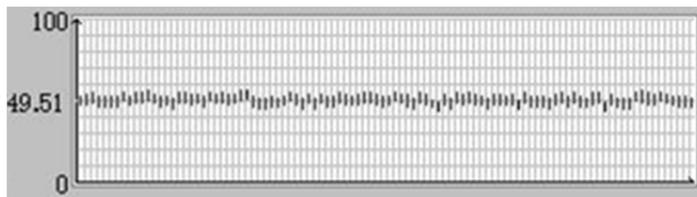
標本数 5



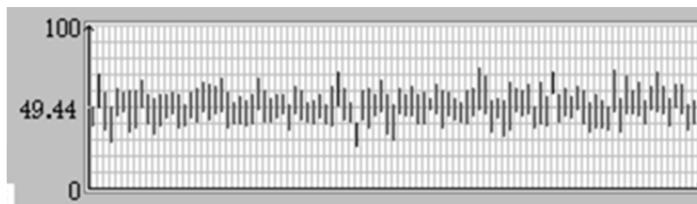
標本数 10



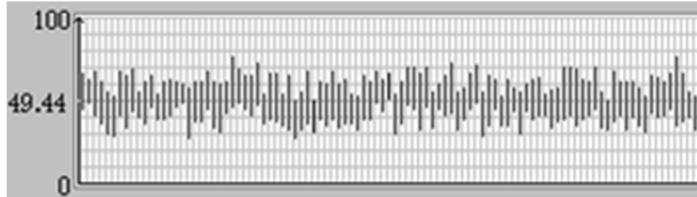
標本数 100



標本数 10  
信頼率 95%



標本数 10  
信頼率 99%



次に実際に信頼区間をつけて標本から母平均を推定してみよう。

① 標本数  $n$  が多い (大標本, 標本数  $n$  が 100 以上) あるいは母標準偏差が既知のとき

標本数  $n$  ( $\geq 100$ ) が十分に多ければ, 母集団の標準偏差は標本の標準偏差と同じだと見なしてよい。したがって, 標本平均  $\bar{x}$ , 標本標準偏差  $s$  とすると,

標準誤差  $SE = \frac{s}{\sqrt{n}}$  となり, エクセルの正規分布に関する関数から, 信頼率  $p$  % における母平均

の区間推定を計算できる。

すなわち標本数  $n$ , 標本平均  $\bar{x}$ , 標本標準偏差  $s$ , 標本の標準誤差  $SE$  とすると, 信頼率  $p$  % における区間推定は,

エクセルの正規分布に関する関数  $NORMINV$  を利用して, 母平均は

$$NORMINV((1-p/100)/2, \bar{x}, SE) \leq \mu \leq NORMINV(1-(1-p/100)/2, \bar{x}, SE)$$

あるいは

$$NORMINV((1-p/100)/2, \bar{x}, s/SQRT(n)) \leq \mu \leq NORMINV(1-(1-p/100)/2, \bar{x}, s/SQRT(n))$$

と計算できる。

あるいは値を代入するだけで計算するように作ったセル (授業用エクセルファイルの「第4回平均区間推定」タブにある) へ代入してもよい。

標本数, 標本平均, 標本標準偏差, 信頼率を入力すると信頼区間を計算するように以下のシートは作ってあります。 適当なセルへコピーして使用してください。			
標本数	100		
標本平均	120		
標本標準偏差	10		
信頼率	95 %		
上限	121.96		
下限	118.04		
±表記	1.959964		

例 100 個のトマトの重さを量ったときに平均 120.00g, 標準偏差 10.00g だった。95%信頼区間と 99%信頼区間をつけて母平均を区間推定せよ。

95%信頼区間

$$118.04 \leq \mu \leq 121.96$$

$$120.00 \pm 1.96$$

99%信頼区間

$$117.42 \leq \mu \leq 122.58$$

$$120.00 \pm 2.58$$

標本数	100
標本平均	120
標本標準偏差	10
信頼率	95 %
上限	121.96
下限	118.04
±表記	1.959964

標本数	100
標本平均	120
標本標準偏差	10
信頼率	99 %
上限	122.5758
下限	117.4242
±表記	2.575829

予習1 100個のトマトの重さを量ったときに平均120.00g, 標準偏差10.00gだった. 90%信頼区間をつけて母平均を区間推定せよ.

予習2 10000個のトマトの重さを量ったときに平均120.00g, 標準偏差10.00gだった. 90%信頼区間をつけて母平均を区間推定せよ.

予習1と予習2の答えを比べれば, 標本数を増やすことによって, 推定精度が上がるのがわかる. 標本数, 信頼率を変えることによって, 母平均の区間推定の範囲が変わっていく様子を観察しよう. 点推定では標本数を変えても点推定値は変わらないから, 精度は点推定では表現されない.

★ 信頼率としては95%あるいは99%を使うことが多いので, 信頼率95%の信頼区間の式  $\bar{x} - 1.960SE \leq \mu \leq \bar{x} + 1.960SE$  および信頼率99%の信頼区間の式  $\bar{x} - 2.576SE \leq \mu \leq \bar{x} + 2.576SE$  を覚えておくとう便利である.

#### 演習問題

S君は大学の研究で近くの森を調査している. そこに生息するある生き物の平均体重を5%以内の精度で推定したい. 5%の精度とはここでは精度(%) = (信頼率95%の信頼区間の1/2) ÷ 平均 × 100%とする (つまり平均 ± 信頼率95%の信頼区間の1/2 と表記するときの±の後に付く数字).

この生き物の体重について, 母平均は250g, 母標準偏差は100gとわかっている. すると何匹, 調査すれば精度を5%以下にできるだろうか?

#### ② 標本数 $n$ が少ないとき (小標本)

このときは標本の標準偏差  $s$  から母集団の標準偏差  $\sigma$  を推定する誤差が大きいため, 正規分布ではなく  $t$  分布で推定する. 詳細は第6回の授業で行う.

#### E. 宿題

宿題は <https://moodle.cerd.shimane-u.ac.jp/moodle/> を見てください.